APPARATUS AND METHOD OF ACHIEVING AND IMPROVING LINEARITY OF AMPLIFIERS AND OTHER NON-LINEAR PATHS

Background to the Invention

This invention relates, in general, to an apparatus and method of achieving a scheme in which signals are linearly amplified, i.e., a linear transfer function is achieved from an input to an output of a non-linear amplifier or path. The invention is applicable to information bearing signals in an analog or digital form and in which a sine wave carrier has its envelope and/or phase varied.

Reference is made to the glossary at the end of the technical description that provides guidance as to the meaning of the various technical terms used in this document.

Summary of Prior Art

With respect to information recovery, generally, amplification of a signal is desirable since data integrity within the signal is enhanced as a consequence of the amplification process providing an increased ability to differentiate the signal from noise or other interfering signals. The amplification of signals is therefore practised in many areas of technology, including (but not limited to) analog and digital telecommunications systems. There are also other applications where precise linear amplification may be required even though the primary purpose is not for information transfer.

Generally, all amplifiers are non-linear because amplitude variations of the input signal (called here the “excitation”) are not faithfully reproduced in the output signal (called here the “response”). In particular, amplifiers of the type most relevant to this invention can be characterised in terms of their response to an excitation by a sine wave of variable amplitude or envelope. The response is actually in two parts, namely an envelope and the phase of the output sine wave. More particularly, it will be appreciated that present-day amplifiers have

an operating characteristic that exhibits reasonable linearity at low operating and power levels, with the degree of linearity being related to component limitations and hence manufacturing costs of the amplifier. The useful operating range of the amplifier, however, only approximates to a linear function over a relatively small range, with the linearity of the characteristic quickly disappearing in the saturation region. In fact, in relation to the linear range, the characteristic is typically bowed and arcuate in nature, while the phase (which should be independent of the excitation) varies in a more complex manner. Unfortunately, any departure from so-called "perfect" response curves can corrupt information carried in the form of amplitude or phase modulation. For example, in relation to an output envelope of a non-linear device and with increasing excitation power, the amplifier response contains both a wanted signal at a lower level than would be the case for a linear amplifier and increasing amounts of distortion that corrupts the integrity of the signal. Not only this, but the distortion can broaden the frequency spectrum, and therefore potentially causes harmful interference in adjacent frequency channels.

Any signal path of a type relevant to this invention can be characterised by the above description, albeit to differing degrees.

A commonly practised method of linear amplification requires the selection of a "quiescent" (operating) point along the amplifier characteristic that is backed-off from the saturation region. Backing-off ensures that excitation signals having a relatively high (initial) level are not distorted by an out-of-range excursion (with respect to the linear range of the amplifier). In other words, backing-off causes even the best linear amplifiers to have restricted performance to ensure operation within the ostensibly linear range. Clearly, the effect of backing-off an amplifier in relation to signal transmission reduces the potential output power and consequently restricts the attenuation that a signal can suffer and yet be successfully recovered.

8th European Conference on Electrotechnics, Stockholm, Sweden, June 13-17, 1988, pp 64-67. LINC forms a constant level carrier whose spectrum is made up of the wanted signal plus other terms that maintain the constant envelope. A second carrier is also formed in which the wanted signal spectrum is inverted but the terms to maintain the envelope, which are the same as for the first envelope, are not. The two carriers form the excitation for two very non-linear (even saturated) amplifiers. The responses are then subtracted by suitable means, leaving the wanted signal and suppressing the unwanted parts of the spectrum. It is important to note that the spectrum of these unwanted components lies in the same frequency range as the signal. It is also of note that the signal envelope must be extracted and processed.

In practice, the problem of non-linearity can be avoided by the use of a constant level signal, e.g. digital modulation using pure phase modulation.

The choice of nominally constant envelope signals is optimum in the sense that amplifier output power can be maximised and non-linearity becomes of second order importance; however, it is not optimum in the sense of bandwidth utilisation. In this respect, it will be appreciated that by allowing envelope as well as phase variations, a given bandwidth can support a higher rate of information transmission.

Unfortunately, all these methods of addressing non-linearity can suffer from one or more of the following disadvantages, namely (i) a reduced maximum output power, (ii) drift (instability), (iii) lack of bandwidth, (iv) an inability to handle a wide range of signal types (i.e., the method is application specific), or (v) high manufacturing costs.

Many communications systems do not require linear amplifiers in the strict sense in that their signals are sensibly constant in amplitude, e.g., the Global System for Mobile communications (GSM). However, imperfections and

processing (e.g., deliberate band limiting) often result in residual envelope fluctuations. If the subsequent amplifier is non-linear, harmful out-of-band emissions can then occur which must be controlled to within stipulated limits regulated by national legislation.

In many situations (including GSM), several signals are usually combined in a common load (or antenna), with such combination by way of several separate amplifiers and suitable couplers. However, if these signals are not sufficiently spaced in frequency, such couplers cannot be lossless and most of the useful power is dissipated; in practice N equal power amplifiers can only deliver $1/N^{th}$ of their individual powers when combined. An alternative to this configuration requires combination of such signals before they act as the excitation for a single high power amplifier, in which case high power amplifier must be very linear (and must, in principle, have the same peak power as all the signals combined).

Another application that requires the use of a linear amplifier is in relation to multi-carrier transmission schemes, such as Discrete Multi-tone (DMT) and Orthogonal Frequency Division Multiplexing (OFDM). Such multi-carrier schemes have been proposed for many types of communications systems, including Digital Audio Broadcasting (DAB), Coded Orthogonal Frequency Division Multiplexing (for the new digital television service (COFDM), and Broadband Wireless Local Area Networks (LANs). The advantage of such schemes is that highly time dispersive channels can provide data transmission rates that are capable of approaching the Nyquist rate for the channel bandwidth.

A highly dispersive channel can arise from a combination of multiple discrete and delayed signals particularly in radio environments, or may be an intrinsic property of the transmission medium (such as a long copper wire pair or an optical fibre) where the group delay is a continuous function of frequency.

Multi-carrier systems are particularly suited to dispersive channels and high bandwidth applications. However, the composite signal envelope produced by DMT, for example, exhibits a high "peak to average ratio (PAR)", which term is sometimes referred to as "peak to mean envelope power ratio (PMEPR)" or the "crest factor". A further disadvantage is that when multiple frequencies are present the amplifier non-linearities cause spurious frequencies called intermodulation products to be formed. To mitigate this a very linear amplifier is required. The conflicting requirement of maximising peak power whilst holding intermodulation to acceptable limits can best be served by an amplifier which is linear right up to the point where the amplifier can deliver no more power (i.e., is saturated)

A further application that requires a linear amplifier of internal path (such as within a receiver) is in the processing of signals, in general, and usually those which have a low amplitude. As but one example, modern receiver design often amplifies and frequency translates a large number of uncorrelated signals occupying a relatively wide frequency band before selection of the one desired. Unless the path is linear, intermodulation products create noise-like interference.

It is desirable, therefore, to develop an apparatus and technique that improves the linearity of a signal path or an amplifier up to its peak power output to enhance signal integrity and amplification performance, generally.

Summary of the Invention

In a first aspect of the present invention there is provided a method of generating a signal principally for use in relation to a non-linear signal path, the method comprising the steps of: combining an information-bearing signal at a first frequency with an idle frequency at a different frequency to generate a combined signal; substantially doubling a phase angle of the combined signal to produce a neoteric signal having a second phase angle; utilising the second

phase angle as phase modulation in the neoteric signal; and ensuring that an envelope of the neoteric signal is substantially constant in level.

The neoteric signal contains a replica of the information-bearing signal and a further component that ensures the envelope is held at a substantially constant level. The frequency difference between the idle frequency and the information-bearing (or narrowband) signal is such that the information-bearing signal can be recovered from the neoteric signal, while other components introduced in the signal generation process are be rejected (preferably as a consequence of being out-of-band and, preferably, of such a form that can be removed by subtraction).

The combined signal may be passed over an internal, non-linear path, while the neoteric signal may be applied to an external non-linear path.

In another aspect of the present invention there is provided a method of generating a neoteric signal comprising the steps of: a) in a first chain: combining a first information-bearing signal at a first frequency with an idle frequency at a different frequency to produce a first zone signal having a phase angle; and constraining an envelope of the first zone signal to a substantially constant level; b) in a second chain: providing a second information-bearing signal with the same phase modulation as the first information-bearing signal but said same phase modulation being in an opposite sense and wherein the second information-bearing signal has a central frequency displaced from an idle frequency by an amount equal to a frequency difference between the first information-bearing signal and its idle frequency but in an opposite sense; combining the second information-bearing signal with its idle frequency to generate a first zone signal having a phase angle; and constraining an envelope of the first zone signal of the second chain to a substantially constant level to form an intermediate signal; and c) taking a frequency difference between the constrained first zone signal of the first chain and the intermediate signal to

generate the neoteric signal having a phase angle substantially twice that of the constrained first zone signal and wherein the step of taking the difference effectively cancels AM to PM conversion introduced by processing in both the first chain and the second chain.

In another aspect of the present invention there is provided a method of generating a neoteric signal comprising the steps of: a) in a first chain: combining a first information-bearing signal at a first frequency with an idle frequency at a different frequency to produce a first zone signal having a phase angle; and constraining an envelope of the first zone signal to a substantially constant level; b) in a second chain: providing a second information-bearing signal with the same phase modulation as the first information-bearing signal but said same phase modulation being in an opposite sense and wherein the second information-bearing signal has a central frequency displaced from an idle frequency by an amount equal to a frequency difference between the first information-bearing signal and its idle frequency but in an opposite sense; combining the second information-bearing signal with its idle frequency to generate a first zone signal having a phase angle; and constraining an envelope of the first zone signal of the second chain to a substantially constant level to form an intermediate signal; substantially doubling the phase angle of the intermediate signal to produce a second intermediate signal; c) taking a frequency sum of the constrained first zone signal and the intermediate signal to generate a third intermediate signal and such as to subtract the respective phase angles while adding respectively associated AM to PM conversions; and d) taking a frequency difference between the third intermediate signal and the second intermediate signal to generate the neoteric signal having a phase angle substantially twice that of the phase angle of the first zone signal and wherein AM to PM conversion introduced by processing in both the first chain and the second chain are effectively cancelled.

In yet another aspect of the present invention there is provided a method of

processing a received signal to recover information, the method comprising the steps of: combining the received signal with an idle frequency having a different frequency to the received signal to produce a first zone signal having a phase angle; constraining a signal envelope associated with the first zone signal to a substantially constant level; substantially doubling the phase angle of the first zone signal to produce a neoteric signal; and selectively filtering the neoteric signal to recover the information.

In another aspect of the present invention there is provided an apparatus for generating a neoteric signal from an information-bearing signal at a first frequency, the apparatus comprising: an idle frequency generator for generating an idle frequency at a different frequency to that of the information-bearing signal; combining means for combining one of the information-bearing signal with the idle frequency to generate a combined signal; means for substantially doubling a phase angle of the combined signal to produce the neoteric signal having a second phase angle and in which the second phase angle acts as phase modulation; and means for ensuring that an envelope of the neoteric signal is substantially constant in level.

In yet another aspect of the present invention there is provided an apparatus for generating a neoteric signal from an incident information-bearing signal at a first frequency, the apparatus comprising: a) an idle frequency generator for generating an idle frequency at a different frequency to that of the information-bearing signal; b) a first chain having: means for combining a first information-bearing signal at a first frequency with the idle frequency to produce a first zone signal having a phase angle providing phase modulation; and means for constraining an envelope of the first zone signal to a substantially constant level; c) a second chain having: means for providing a second information-bearing signal with the same phase modulation as the first information-bearing signal but said same phase modulation being in an opposite sense and wherein the second information-bearing signal has a central frequency displaced from an

idle frequency by an amount equal to a frequency difference between the first information-bearing signal and its idle frequency but in an opposite sense; means for combining the second information-bearing signal with an idle frequency to generate a first zone signal having a phase angle; means for constraining an envelope of the first zone signal of the second chain to a substantially constant level to form an intermediate signal; and d) means for taking a frequency difference between the constrained first zone signal of the first chain and the intermediate signal to generate the neoteric signal having a phase angle substantially twice that of the constrained first zone signal and wherein the step of taking the difference effectively cancels AM to PM conversion introduced by processing in both the first chain and the second chain.

In still yet another aspect of the present invention there is provided apparatus for generating a neoteric signal from an incident information-bearing signal at a first frequency, the apparatus comprising: a) a first chain having: means for combining a first information-bearing signal at a first frequency with an idle frequency at a different frequency to produce a first zone signal having a phase angle; and means for constraining an envelope of the first zone signal to a substantially constant level; b) a second chain having: means for providing a second information-bearing signal with the same phase modulation as the first information-bearing signal but said same phase modulation being in an opposite sense and wherein the second information-bearing signal has a central frequency displaced from an idle frequency by an amount equal to a frequency difference between the first information-bearing signal and its idle frequency but in an opposite sense; means for combining the second information-bearing signal with an idle frequency to generate a first zone signal having a phase angle; and means for constraining an envelope of the first zone signal of the second chain to a substantially constant level to form an intermediate signal; substantially doubling the phase angle of the intermediate signal to produce a second intermediate signal; c) means for taking a frequency sum of the

constrained first zone signal and the intermediate signal to generate a third intermediate signal and such as to subtract the respective phase angles while adding respectively associated AM to PM conversions; and d) means for taking a frequency difference between the third intermediate signal and the second intermediate signal to generate the neoteric signal having a phase angle substantially twice that of the phase angle of the first zone signal and wherein AM to PM conversion introduced by processing in both the first chain and the second chain are effectively cancelled.

In yet still a further aspect of the present invention there is provided a receiver arranged to recover information from incident information-bearing signals, the receiver comprising: means for combining incident information-bearing signals with an idle frequency having a different frequency to the incident information-bearing signals to produce a first zone signal having a phase angle; a limiter for constraining a signal envelope associated with the first zone signal to a substantially constant level; means for substantially doubling the phase angle of the first zone signal to produce a neoteric signal; and a filter for selectively filtering the neoteric signal to recover information.

Advantageously, therefore, the present invention provides a signal processing scheme that significantly improves the linear response of amplifier circuitry (or any other path or internal path as described) by preventing distortion of the wanted signal as it progresses through the amplifier and at the same time allows the full peak power of the amplifier to be diverted into the peak power of the wanted signal. Furthermore, the present invention can be implemented at minimal cost for the signal processing by means of analog methods, or in certain cases by a dedicated digital signal processor, which processor may be already provisioned within existing communication device architectures. The present invention therefore improves the efficiency of information transfer a) by reducing distortion; and b) by an associated increase in the allowable signal attenuation of a transmission medium. Furthermore, in relation to amplifier arrangements

that make use of the properties of the signal processing scheme of the current invention, such amplifier arrangements generally operate more efficiently in relation to both power consumption and performance since the increased linearity attained by the information bearing signal allows an amplifier to be driven in optimal fashion. Indeed, the scheme of the present invention supports linear amplification of signals without a significant penalty in respect of the maximum output capability of the peak output of the amplifier.

Brief Description of the Drawings

Exemplary embodiments of the present invention will now be described with reference to the accompanying drawings, in which:

FIGs. 1(a) and 1(b) contrast response characteristics of perfect and non-linear amplifiers;

FIG. 2 is a mixed flow and operational diagram illustrating the principal signal processing steps involved in the generation of a neoteric signal according to a preferred embodiment of the present invention;

FIG. 3 shows, in vector form, both the generation (FIG. 3(a)) of the neoteric signal according to a preferred embodiment and the geometric properties (FIG. 3(b)) of the neoteric signal;

FIG. 4 contrasts a spectrum of the neoteric signal of the preferred embodiment against the spectra of its individual (but modulated and unprocessed) constituent components;

FIGs. 5 to 7 illustrate alternative signal processing chains according to various preferred embodiments of the present invention;

FIG. 8 resembles the signal processing chains of FIGs. 5 or 6, but further includes preferred mechanisms that compensate for small distortions that can occur anywhere along the signal path;

FIG. 9 is yet another embodiment of the present invention, albeit that this embodiment is suitable for use on a path, such as an optical modulator, in which a ratio of the maximum and minimum frequencies of the information-bearing signal is typically large;

FIG. 10 is a further embodiment of the present invention having an architecture that removes amplitude to phase modulation conversion products occurring within the signal processing;

FIG. 11 is yet another embodiment that controls overload of the neoteric signal;

FIG. 12 shows a preferred embodiment of the invention to illustrate that the internal path contained within the processing to obtain the neoteric signal can be linearized as well as a path using the neoteric signal itself. The example chosen is a complete receiver which makes use of one or more of the embodiments of FIGs. 5-11; and

FIGs. 13 and 14 show preferred architectures that are necessary to recover the wanted signal from the neoteric signal after its passage through an amplifier.

Detailed Description of a Preferred Embodiment

Reference is briefly made to FIGs. 1(a) and 1(b) that contrasts the operating characteristics of non-linear and linear amplifiers in relation to a varying excitation signal. As will be appreciated, the useful operating range of a non-linear amplifier is limited to a region below the saturation point, while the overall characteristic only approximates to a linear function over a relatively small range. In other words, the linearity of the characteristic quickly disappearing in the saturation region, as shown in FIG. 1(a). FIG. 1(b) represents a phenomenon called amplitude modulation to phase modulation (AM to PM) conversion. Any signal path of the type relevant to the invention can be characterised by such curves; only the shape of the curves changes.

Before undertaking a detailed description of a signal processing mechanism of a preferred embodiment of the present invention, the basic inventive premise of the present embodiment will be summarised. The preferred embodiment of the present invention can be considered in relation to an electrical circuit, although the technique is functionally the same for an optical device to the extent that light can be regarded as simply an extension of the electromagnetic spectrum, or indeed, any other device that uses narrowband signals. An input signal to be

amplified is converted (i.e., pre-processed by the addition of out-of-band components) into the neoteric signal having a constant envelope (or amplitude). When the input signal is zero the signal merely consists of an unwanted component and when the input signal is at the maximum possible level, the neoteric signal consists of the wanted signal alone (this being a replica of the input signal) with no unwanted component. At all intermediate levels there is a mixture of the wanted signal and the unwanted component but the neoteric signal envelope nonetheless remains constant. As such, the processing of the preferred embodiment of the present invention expands the bandwidth of the neoteric signal with respect to the input signal. The wanted components that cause the expansion can be arranged to be out of the frequency band of the information-bearing signal and can be filtered out. Additionally, residual components can be removed by a process involving subtraction.

One preferred embodiment of the present invention principally operates to provide a method of effectively improving the linearity of narrowband amplifiers by exploiting the properties acquired from a constant amplitude excitation signal. In these respects, the present invention enables the amplifier to support a response signal that has the same spectrum as the excitation signal. The neoteric signal is thus suitably structured to be the excitation to an amplifier; with the constant amplitude of the neoteric signal providing an ability to drive the amplifier, if required, to its maximum permitted level. Thus, by using one of the preferred embodiments of the present invention, linearity for the wanted signal is achieved over the whole operating range of the amplifier. The response signal, containing an amplified replica of the wanted signal, can now be processed to extract the wanted signal and reject the unwanted components.

A second preferred embodiment of the present invention principally operates to provide a method of linearizing a signal path by exploiting the properties acquired as an integral part of the generation of the neoteric signal. The generation involves non-linear processing (i.e., a limiting of the envelope of the

sum of an input signal and an idle frequency f_i to generate the neoteric signal) which, however, has no effect on the wanted signal contained within the neoteric signal. Therefore, by making a non-linear internal path, e.g., a small signal amplifier, part of the processing to obtain the neoteric signal, the wanted signal is again a replica of the input signal. Moreover, it will be appreciated that where the so-called internal path is approximately linear, the sum of the input signal and the idle frequency f_i, while having a somewhat expanded bandwidth relative to the input signal (due to so-called mild non-linearities) will, however, occupy much less bandwidth than that of the neoteric signal.

More particularly, the preferred embodiments of the present invention convert a narrowband input signal containing any mixture of envelope and phase variations into a new signal having a constant amplitude made up of a wanted signal and an added unwanted component in an inventive way. The added component can be removed by filters or by a combination of filters and subtractions.

The present invention can be explained, principally, in three ways, namely: (i) a flow diagram; (ii) in relation to vector diagrams; and (iii) in relation to the frequency domain.

FIG. 2 shows a mixed flow and operational diagram illustrating the principal signal processing steps involved in the generation of a neoteric signal according to a preferred embodiment of the present invention. The neoteric signal of the preferred embodiment can then be used to provide an excitation signal for an amplifier, or can be processed to extract the wanted signal where the internal path is of interest.

A signal of interest (i.e. the information-bearing signal) 50 that is to be amplified is added 52 to a fixed level idle frequency f_i (reference numeral 54) that is generated 56 locally. The level of the idle frequency f_i is fixed so that it is always

higher than or equal to the peak envelope of the original signal 50. However, in a preferred embodiment, the level of the idle frequency f_i can be varied 58 for special applications (to be described later). A combination of the two sine waves of the original signal 50 and the idle frequency f_i can be considered as a carrier 60. The carrier has simultaneous envelope and phase modulation. Moreover, the carrier 60 can, if required, be applied 62 to a so-called internal path that is non-linear; this optional step is shown by the dotted nature of the function box in FIG. 2.

From a practical standpoint, only the phase modulation α, a definition of which is found later when describing the vector diagram of a preferred embodiment of the present invention, is of interest in relation to the present invention.

According to the present invention, the neoteric signal 64 is now formed by both altering 65 the phase modulation to substantially 2α and by constraining 66 the signal envelope. With respect to the step of constraining the signal envelope 66, this can be applied at any one of the preceding stages, namely after the stage of: i) applying 62 the carrier 60 to a non-linear internal path; or ii) at any stage after forming 65 the phase modulation 2α. Further, as part of the processing to form 2α, the information-bearing signal, the idle frequency, the constrained combination of these two signals and the neoteric signal may be combined in different ways to give a phase modulation that adds beneficial distortion to the information-bearing signal.

The wanted signal 67 can, at this point, be extracted 68 directly to provide output 70. Extraction of the wanted signal is achieved by applying the neoteric signal 64 to a suitably shaped filter (not shown) arranged to reject undesired spectral components occurring outside a stipulated bandwidth centred about the wanted signal. Indeed, the extraction of the wanted signal 67 is made easy because the wanted signal and the unwanted component are arranged to fall in different parts of the frequency spectrum. Alternatively, in certain embodiments

a part of the unwanted component can fall within the wanted signal spectrum but can be removed be subtraction.

Alternatively, it may be desirable to continue processing of the neoteric signal 64 in order to form the excitation 71 for an amplifier (which, from the definition, can also be considered as a path in its own right). However, since the neoteric signal 64 has a wide bandwidth, its spectrum can first be optionally controlled 72 to match the characteristic of the amplifier (as represented by block 74). The excitation 71 preferably drives the amplifier 74 to its maximum permitted power. The amplifier response 76, which is still substantially the neoteric signal, has the wanted signal extracted 68 while the unwanted components are rejected, as previously described. The wanted signal 67 is then made available at a second output 78 for further signal processing (as desired), with the wanted signal 67 undistorted and having a peak power substantially equal to the maximum permitted power of the amplifier.

FIG. 3 shows, in vector form, both the generation (FIG. 3(a)) of the neoteric signal according to a preferred embodiment and the geometric properties (FIG. 3(b)) of the neoteric signal.

In FIG. 3(a), vector OA (of unit length for convenience) represents a constant level sine wave of the idle frequency f_i that rotates at an angular rate $2\pi f_i$. Vector AB of length "r" (where r<1) and phase modulation ϕ represents the original signal 50 that is added to the idle frequency: the original signal 50 rotates relative to OA at an angular rate $p=2\pi(f_s-f_i)$, where f_s is the frequency of the original signal 50. The resulting vector OB of length "m" has a locus that lies on the circle of radius "r", as shown, and an angle α relative to OA. It will be understood that FIG. 3(a) is a snapshot taken for a fixed value of "r".

For convenience in the explanation, the vector OB is now assumed constrained (by functional block 66 of FIG. 2) to lie on the unit circle but its angle α remains

unaltered. The resulting vector OB' will describe an arc as vector AB rotates and the maximum swing of α will be ±90°; this occurs for α value of r=1. Such a signal can be obtained directly from a limiter. It is noted that FIG. 3(a) is only used for convenience in defining the angle α. More specifically,

$$\arctan(\alpha) = \frac{r\sin(pt+\phi)}{1+r\cos(pt+\phi)} \tag{1a}$$

and

$$m = \sqrt{\{1+2r\cos(pt+\phi)+r^2\}} \tag{1b}$$

FIG. 3(b) shows the vector diagram for the neoteric signal 64 (vector OC) of the preferred embodiment of the present invention, which neoteric signal 64 is obtained from substantially doubling in angle of the vector OB' relative to the reference vector OA. The neoteric signal 64 can be obtained by many means but typically is achieved by passing the signal represented by the vector OB' through a frequency doubling device to form a second zone signal. However, any method which forms a sensibly constant amplitude signal with a phase variation of substantially 2α will achieve the same result. It is noted that there is a trivial shift in frequency associated with generation of the second zone signal, but this can be corrected for elsewhere within the signal processing path. A substantial doubling (and preferably exact doubling) of the angle α gives a vector OC that has a constant amplitude and whose locus lies on the full dotted circle 90 shown in FIG. 3(b). The maximum swing of vector OC will be 2α=±180° when r=1.

The vector OC can also be considered to be made up of two components, namely OD and DC. To show that DC is an exact replica of the original signal 50, the dotted construction lines, namely an extension of OD to B' and B, and an arc of radius "r" centred on A have been drawn on FIG. 3(b). Points D and B are

unaltered. The resulting vector OB' will describe an arc as vector AB rotates and the maximum swing of α will be ±90°; this occurs for α value of r=1. Such a signal can be obtained directly from a limiter. It is noted that FIG. 3(a) is only used for convenience in defining the angle α. More specifically,

$$\arctan(\alpha) = \frac{r\sin(pt+\phi)}{1+r\cos(pt+\phi)} \tag{1a}$$

and

$$m = \sqrt{\{1+2r\cos(pt+\phi)+r^2\}} \tag{1b}$$

FIG. 3(b) shows the vector diagram for the neoteric signal 64 (vector OC) of the preferred embodiment of the present invention, which neoteric signal 64 is obtained from substantially doubling in angle of the vector OB' relative to the reference vector OA. The neoteric signal 64 can be obtained by many means but typically is achieved by passing the signal represented by the vector OB' through a frequency doubling device to form a second zone signal. However, any method which forms a sensibly constant amplitude signal with a phase variation of substantially 2α will achieve the same result. It is noted that there is a trivial shift in frequency associated with generation of the second zone signal, but this can be corrected for elsewhere within the signal processing path. A substantial doubling (and preferably exact doubling) of the angle α gives a vector OC that has a constant amplitude and whose locus lies on the full dotted circle 90 shown in FIG. 3(b). The maximum swing of vector OC will be $2\alpha=\pm180°$ when r=1.

The vector OC can also be considered to be made up of two components, namely OD and DC. To show that DC is an exact replica of the original signal 50, the dotted construction lines, namely an extension of OD to B' and B, and an arc of radius "r" centred on A have been drawn on FIG. 3(b). Points D and B are

called the two intersections of the arc of radius r with vector OB, while the arc and the extended vector OB are, of course, replicas from FIG.3(a). Construction of an arbitrary line Da parallel with OA then assists in showing that vector DC is parallel with vector AB. It is also noted that AD=AB=r.

Vector OC can now be broken down into two components, although at this point the breakdown is arbitrary. Consider the triangles ODC and ODA. In these OC=OA=1, OD is common and the angle COD=AOD=α. The triangles are therefore identical (because two sides and the included angle are the same) and it follows that DC=AD=r. Consider the triangles AOB and COB. In these, OC=OA=1, while OB is common and angle COB=AOB=α. Hence, these triangles are also identical and therefore CB=AD=r. ABCD is then a rhombus and hence DC and AB are parallel. Therefore, not only is the vector DC equal to r in length it also has the same angle CDa=pt+ϕ relative to the reference OA; it is, in fact, an exact copy of the original signal.

It will be appreciated that if the angle COA may not be exactly equal to 2α either because of design approximations or spurious added phase (e.g. due to amplitude to phase conversion in the processing path). As such, the vector DC will not be an exact replica of AB but as long as the differences are small they can be generally ignored or can be used beneficially.

It will be appreciated that, as the wanted signal vector DC varies, the vector OD has a variable length and phase α relative to the reference OA. As such, vector OD represents an unwanted component. By application of the laws of triangles it can then be shown that OD=$(1-r^2)/m$ since sides OC and DC and angle COD are known.

It will be shown later that the two vectors DC and OD represent two components of the neoteric signal, and that by a suitable choice of the frequency difference p between the original signal and the idle frequency f_i that the unwanted

component can be made to fall outside the frequency band of the wanted signal or that a substantial part of it can be removed by subtraction. Consequently, the unwanted component OD can be rejected by suitably shaped and frequency centred filters.

While FIG. 3 is drawn with a small angle $pt+\phi$, the diagrams are valid for all values of $pt+\phi$ (although when the angle OBA becomes larger than 90° it is necessary to use the projection of AB onto the unit circle).

FIG. 4 contrasts a spectrum of the neoteric signal of the preferred embodiment (FIG. 4(b) against the spectrum of its individual (but modulated and unprocessed) constituent components (FIG. 4(a)). Specifically, FIG. 4(a) shows a frequency domain representation of vectors OA and AB. In FIG. 4(a), signal f_s is the original signal 50 having a spectrum of width "B" caused by applied modulation. Meanwhile, the idle frequency f_i (identified by numeral 54), is displaced from the input signal by a frequency P hertz, where $P=p/2\pi$ (p is the radian frequency). For convenience, P has been taken to be negative (i.e. $f_s<f_i$) relative to the idle frequency.

Now, according to the preferred embodiment of the present invention, FIG. 4(b) illustrates a spectrum 104 of the neoteric signal 64 (i.e. the vector OC of FIG 3(b) containing a phase variation substantially twice that of vector OB or OB' of FIG. 3(a)). Later, it will be shown mathematically that the spectrum 104 contains the wanted signal 67 (which is a replica of the original signal 50) plus a series of unwanted components 106-112 that extend indefinitely. The series of unwanted components, in spectral totality, represent vector OD of FIG. 3(b)), which unwanted components 106-112,... can be arranged by selection of P to have no effect on the wanted signal 67 since preferably no overlapping occurs (in the preferred embodiment) between the wanted signal spectrum 67 and certain unwanted components (namely 106, 110, 114,...) can be removed by subtraction. It will, however, be observed that the unwanted components 108-

112... each have separate but overlapping spectra, while the span of each unwanted component increases as 2B, 3B, 4B, 5B and so on. Furthermore, each unwanted component is nominally centred on an integral frequency spacing at *P* hertz. For a complete understanding it is noted that in the absence of any modulation on the original signal 50 the spectra of the unwanted components become line spectral at integral frequency separations of *P* hertz.

By inspection, there is a separation of *P*-3B/2 between the closest edge of the wanted signal 67 and the unwanted component 106. There is a separation of 2*P*-2B between the closest edge of the wanted signal 67 and the unwanted component 108. The exact spectrum of the unwanted components cannot be predetermined because it is dependent upon the modulation applied to the original signal 50, although it can be shown that unwanted spectrum cannot impinge on the wanted signal provided that phase doubling to 2α occurs and the frequency separation between the original signal 50 and the idle frequency is sufficient. Indeed, it is the separation which enables the wanted signal 67 to be extracted by means of a band-pass filter having an appropriated shaped filter response curve 120 (as shown on FIG. 4(b)). Alternatively, in certain embodiments, unwanted components 106, 110, 114,... may be removed by a technique of subtraction and the component 106 can fall within the passband of the wanted signal. As such, a separation of less than 3B/2 can be tolerated provided that unwanted components emanating from the idle frequency f_i are substantially cancelled.

In relation to FIG. 4(b), it is to be noted that the spectrum 104 is essentially one-sided. However, if *P* is chosen to be positive (i.e. $f_s > f_i$) then the spectrum 104 is mirrored about f_i and the various unwanted spectral components 106-112... fall at a lower frequency than f_i.

To demonstrate that the neoteric signal 64 obtained from a combination of the original signal f_s and a frequency displaced idle frequency f_i contains a term that

is a replica of the wanted signal 67, one can also consider that any narrowband signal can be expressed in a complex form $ae^{j(\omega_i t+\phi)}$, where a is the envelope, ω_i is the angular frequency in radians per second and ϕ is any phase modulation.

Therefore, in relation to a combined input of the wanted signal f_s (having a time variable envelope (or amplitude) "a") and the idle frequency f_i (having an arbitrary but constant amplitude "b"), the combined input v can be expressed as:

$$v = be^{j\omega_i t} + ae^{j(\omega_s t+\phi)}$$

In this expression, ϕ is a time variable phase and b is assumed to be greater than the peak of a, while ω_i and ω_s are, respectively, the angular frequency for the idle frequency f_i and the wanted signal f_s. In the event that a is greater than b, then the wanted signal and the idle tone are interchanged in what follows. However, it is to be noted that this condition represents overload and is, in general, to be avoided.

Hard limiting the combined signal v generates a signal $(v+v^*)^{1/2}$, since hard limiting is effectively a division of the combined signal by its modulus (i.e., $(v.v^*)^{1/2}$ and where * denotes the complex conjugate). The hard limited signal can then be multiplied by itself (i.e., squared) to yield a second zone v_2 (according to a preferred embodiment of the present invention):

$$v_2 = \left[\frac{1+(a/b)e^{j(pt+\phi)}}{1+(a/b)e^{-j(pt+\phi)}}\right]e^{2j\omega_i t}$$

which is the neoteric signal in complex form and where $p = (\omega_i - \omega_s)$ and in which v_2 now has a constant level of unity.

By putting a/b =r and noting that the down conversion of v_2 simply requires a multiplication by $e^{-j\omega_i t}$, v_2 can be re-arranged in a form yielding the neoteric signal at the correct frequency ω_i (which, it is to be noted, is not the signal

frequency ω_s). Then,

$$v_2 = \left\{ re^{j(pt+\phi)} + \left(\frac{1-r^2}{1+re^{-j(pt+\phi)}} \right) \right\} e^{j\omega_i t}$$

$$v_2 = \left\{ re^{j(pt+\phi)} + \left(\frac{(1-r^2)e^{j\alpha}}{m} \right) \right\} e^{j\omega_i t}$$

The first term in the second expression immediately above is the wanted signal (vector DC of FIG. 3(b)) which is actually at its original frequency $\omega_s=\omega_i+p$, while the second term is the unwanted component vector OD. Finally, by taking the first arrangement and expanding the denominator by means of the binomial theorem, the neoteric signal is obtained in the convenient form:

$$v_2 = re^{j(\omega_s t+\phi)} + (1-r^2)[e^{j\omega_i t} - re^{j(\omega_i t-pt-\phi)} + r^2 e^{j(\omega_i t-2pt-2\phi)} - \ldots] \qquad (2)$$

As will be appreciated, the first term in the expression is just the wanted signal at f_s divided by the constant *b*. The remaining terms are all weighted by $(1-r^2)$ and form the unwanted but necessary components 106-112,...of FIG. 4(b) to keep the envelope of v_2 constant. In fact, the general term for each amplitude of each unwanted component is $r^n(1-r^2)$ (for n=0 ,1, 2, 3...). As such, each successive higher order unwanted component 106-112... has a diminishing amplitude, while at r=1 all disappear.

In the preceding expression for the second zone, the first term $re^{j(\omega_s t+\phi)}$ has, of course, the same spectrum as the wanted signal f_s. For the remaining terms that are functions of $(1-r^2)(re^{j\phi})^n$ it can be shown via the convolution theorem that their spectra can never be wider than B(n+2), where B is the total signal bandwidth of the wanted signal f_s. These spectra are centred on f_i+nP, as previously described and shown in FIG. 4(b).

Attention is now turned to suitable mechanisms and apparatus by which the

neoteric signal of the preferred embodiment can be successfully generated. While many methods exist, as will be readily appreciated, some of the principal methods and structural architectures will now be described in two parts, namely: i) the generation of the neoteric signal and its application to an amplifier; and ii) then the recovery of the wanted signal from an output of an amplifier. The so-called "internal path" will also be included.

Referring to FIG. 5, which is probably one of the simplest and most convenient methods and structures, a time varying input signal f_s and the constant amplitude idle frequency f_i, produced by generator 130, are combined in a summation unit 132 to form a carrier 60 with a variable amplitude and phase, as already described in relation to vector OB of FIG. 3(a). An optional internal path 134 may be present between the summation unit 132 and a limiter 136 that constrains the carrier 60 to have a constant envelope (as represented by vector OB' in FIG. 3(a)). The constant envelope signal OB' is then applied to a frequency doubler 138 that generates the neoteric signal, albeit centred on a frequency of $2f_i$, typically by passing through a wide range of readily appreciated devices arranged to generate a second harmonic (or second zone). Preferably, the second zone is isolated by a filter 140.

Optionally, a frequency down converter 142 (responsive to the neoteric signal at $2f_i$ and coupled to the idle frequency generator 130) uses the idle frequency f_i to produce a final neoteric signal 64 in which the wanted signal is coherent with the original signal 50 (i.e. the time varying input signal f_s of FIG. 5). In the case where the internal path is used, it may be desirable to have a second idle frequency generator located locally with respect to the down converter 142 and the wanted signal recovery point. In this way, remote operation is possible although coherence with the time varying original signal may be lost.

Following down-conversion, the neoteric signal 64 is then able to form the excitation for an amplifier or it can be applied immediately to filter 144 to extract

the wanted signal after passage through the internal path 134.

Although not essential to the operation of the present invention, embodiments may include optional filter-limiter combinations (repeated n-fold as necessary, and omitted from the drawings for the sake of clarity) to control and minimise the spectrum of the neoteric signal. The filter-limiter combinations can be inserted at nodes 146-150. The limiter removes amplitude variations caused by the filters and ensures that the process can be repeated as desired. It will also be appreciated that the neoteric signal exists in two frequency regions, namely centred on $2f_i$ and f_i, and as such any filtering and limiting to enhance the spectrum can be achieved equally for the neoteric signal in either frequency region.

FIG. 6 shows another embodiment for generating the neoteric signal of the present invention. A summation unit 132 adds together the original signal f_s and the idle frequency f_i to generate a carrier 60 with a variable envelope and phase. A frequency doubler 158, realised by an even order device (such as an amplifier with enhanced even order harmonic distortion or an up-converter in which the carrier provides both the input and the local oscillation), then produces a second zone which, however, may still contain considerable envelope fluctuations. The second zone is isolated by a suitable filter 160 and the envelope is constrained by limiter 136, at which point the neoteric signal has been formed. As in FIG. 5, n-fold filtering and limiting can be included within the chain; specifically at nodes 162-164 located on either side of an optional frequency down-converter 142. Again, n-fold filtering minimises the spectrum of the neoteric signal, as will be appreciated. The frequency down converter 142, using the idle frequency f_i, generates a final neoteric signal 64 in which the wanted signal is coherent with the input signal. The final neoteric signal 64 is therefore suitable for application to an amplifier, as desired. FIG. 6 further illustrates that limiting can take place anywhere in the signal processing, as already indicated in FIG. 2 and also shows that an internal path 134, as described above, can be incorporated

between summation unit 132 and frequency doubler 158.

It will be further appreciated that FIGs. 5 and 6 represent but two of many possible implementations. Essentially, any architecture that can generate a second zone (whether that zone falls at the correct frequency or not) can, by suitable frequency conversion and constraint (limiting) of the envelope, form the neoteric signal containing phase modulation 2α. By way of example, a third order intermodulation product containing a 2,1th zone can achieve the desired neoteric signal. Furthermore, any architecture that can generate a carrier containing substantially the phase modulation 2α represents a suitable implementation.

The reason for controlling the spectrum of the neoteric signal in FIGs. 5 and 6 is to achieve a practical realisation. Specifically, if the amplifier to which the neoteric signal is applied has a relatively narrow bandwidth then there can be undesired effects within the amplifier. However, constraining the spectrum of the neoteric signal will itself create distortion of the wanted signal and therefore an acceptable and sensible balance between these extremes represents a pragmatic approach, as will be readily appreciated.

As will be understood, the processing to obtain the neoteric signal containing phase modulation 2α can be provided for by an analog or digital device (such as a digital signal processor, DSP). The use of such techniques is preferably when the desired phase and envelope modulation is known or can be obtained by suitable digitization techniques. In this respect, FIG. 7 shows a digital realisation by a processing chain 180 in which amplitude modulation a(t) and the phase modulation ϕ(t) are incident to and digitized (182) by the chain 180, or are already presented to the chain 180 in the form of a digital data stream. A computational block 184 then determines the angle α based on the expression given in eqn. 1 above.

While an angle 2α is required to form the neoteric signal in which the wanted signal does not suffer any significant distortion, the actual value can be chosen to be $k\alpha'$ (or proportional to a function of α, e.g. $k\text{fn}(\alpha)$). Taking the sine and cosine of $k\alpha'$ (in block 186) gives two signals that can be converted back to analog form (in processing blocks 188-190). Specifically, analog cosine and sine signals (192-194) are applied to respective balanced modulator circuits 196-198 that respectively receive as their second inputs time varying cosine and sine components of the idle frequency f_i. The balanced modulator circuits 196-198 have the effect of generating first zone signals, while application of the first zone signals to a summation unit 200 (and limiting of its envelope) generates the neoteric signal having a form $\cos(2\pi f_i t+k\text{fn}(\alpha))$. It is equally possible to make k substantially equal to unity in which case the resulting signal can have its phase angle doubled by analog methods thereby avoiding AM to PM conversion. Further, such choice of k and α can also correct for small distortions that could occur elsewhere, e.g. in the ultimate amplifier due to a bandwidth constraint or actually within a propagation path.

The foregoing explanations concerning the operation and construction of the linearized signal of the preferred embodiment of the present invention have largely assumed that processing is perfect. In reality, however, all active circuits are non-linear in nature and, as such, the signal processing, e.g., limiting, can give rise to undesired distortion of the wanted signal. Beneficially, any non-linearities that affect the signal envelope (i.e., after the addition of the idle frequency in FIGs. 5 or 6) are unimportant, since the processing of the preferred embodiments is, in any case, highly non-linear (and the envelope is removed entirely by a limiter at some stage in the processing). Indeed, it is this feature that makes possible the incorporation of the internal non-linear path within the processing path. However, if a signal has its phase changed by the envelope there will be, in general, AM to PM conversion. The wanted signal also suffers this spurious modulation and there no longer exists absolute linearity between the input and the wanted signal. Another, but controllable, cause of non-linearity

is excessive band limiting of the neoteric signal.

Fortunately, it is possible to correct for any inherent non-linearity in any of the realisations (such as in FIGs. 5 or 6) used to generate the neoteric signal, including any that are present before processing takes place or that are caused to the wanted signal contained within the neoteric signal in its subsequent path. Another method contemplated by a preferred embodiment can be summarised as deliberate distortion of the neoteric signal. In principle, control of the amplitude and phase of this distortion can then be made to cancel most of the unwanted distortion.

Compensation for non-linear distortion can be achieved using the numerous but alternative structures shown in FIG. 8 (in which the presence of repeated limiter-filter combinations have been omitted for the sake of clarity). In essence, at least one of the original signal, the idle frequency or the constrained first zone signal is added at one or more points in the processing chain of FIG. 8 (which can include an internal path 134). FIG. 8 is generally based on FIG. 5 (but it will be appreciated that it can be based on any embodiment including but not limited to FIG. 6 and FIG. 9 to be described later) and principally contains a series combination of one or more summation units 220, limiters 222, a frequency doubler 224 and down conversion circuitry 226. An idle frequency generator 130 provides the idle frequency f_i to a first summation unit 220 that is further responsive to the input signal f_s. In FIG. 8, both the summation unit 220 and the limiter 222 are shown in a plurality of alternate positions that define the various alternative routes for the input signal f_s, the constrained first zone signal and the idle frequency f_i. After each summation unit there is preferably a limiter 222 to remove amplitude variations.

In addition to the function blocks of FIG. 5, FIG. 8 also includes amplitude and phase control circuitry 230-238 for the input (original) signal f_s and the output 104 from the first limiter 222. Time delays, as shown, may be incorporated in the

Therefore, the neoteric signal can lie in a frequency band in which the spectrum of the wanted signal 67 is just clear of zero frequency while still retaining narrowband characteristics. A neoteric signal of a low frequency form can be useful for the transmission of information by, for example, cable distribution networks containing low pass amplifiers or optical modulators and demodulators. FIG. 9 is an embodiment of the present invention for generating and using such a low frequency signal, which embodiment is based on FIG. 5 and in which elements such as filters have been omitted for the sake of simplicity. The only significant difference between FIG. 5 and FIG. 9 is that the signal is first converted to a suitable high frequency for the processing to form the neoteric signal and is then down converted to the correct frequency. The time varying input signal f_s and constant amplitude idle frequency f_i, provided by generator 130, are combined in summation unit 132 to form a carrier 60 with variable amplitude and phase. An optional internal path 134 may be present between the summation unit 132 and up-converter 250 which uses a frequency f_a from frequency generator 252. In essence, the processing is similar to FIG. 5. After combining the signal 50 with the idle frequency f_i, the result is up converted 250 using a reference f_a and is limited 136. Frequency doubling then generates the neoteric signal but one centred on $2f_a+2f_i$. Down converting 258 with reference $2f_a$ followed by further down conversion (with reference f_i in block 142) then yields the required neoteric signal containing the wanted signal coherent with the input signal f_s, which neoteric signal 64 can be applied to a low pass amplifier, optical modulator or other non-linear path. The wanted signal 67 can also be extracted by filter 144.

Where internal path 134 is used, it may be necessary to use a second idle frequency generator located proximate to converter 142. Furthermore, carrier 60, after passage through the internal path 134, must meet the definition of a narrowband signal.

The location of the limiter, frequency doubler and down conversion block 142

paths containing the input signal f_s and the constrained first zone signal output from the first limiter 222, which time delays can serve for the purposes of equalisation. For the idle frequency f_i, the amplitude and phase control circuitry 230-232 can be located in any one or both of the paths between idle frequency generator 130 and summation units 220, while amplitude and phase control circuitry 234-236 associated with the input signal f_s can be located in any one or both of the paths between the input and summation units 220. Further amplitude and control circuitry delivers an output from first limiter 222 to the third summation unit 220 alone via amplitude and phase control circuitry 238. It will be appreciated that FIG. 8 is a composite diagram and that utilisation of only one of the possible paths may be necessary to achieve compensation. These methods involving phase and amplitude modification can also compensate for any non-linear effects anywhere in the neoteric signal path, e.g., caused by band limiting, as well as compensating for non-linear effects on the so-called internal path as well as any information-bearing signal path itself.

FIG. 8 is one of many embodiments whereby deliberate distortion is generated to modify undesired distortion that can occur anywhere in an extended wanted signal path (be it an internal path or otherwise).

In the case where the neoteric signal is generated digitally, there is an increased flexibility as previously discussed. The nominal phase modulation angle 2α can be made $k\alpha'$, where in the simplest case k is a constant (approximately equal to 2). Of course, k could also be made to be some complex function of the modulation, if desired. The effect of having a deviant phase angle in the neoteric signal will be to generate small distortions as described for FIG. 8 above and as discussed earlier for the construction of vector OC in FIG. 3(b); these beneficial distortions may be used to compensate for distortion occurring elsewhere, as already described.

It will be appreciated that the spectrum of the neoteric signal is one-sided.

can, of course, be elsewhere within FIG. 9, as previously indicated in relation to the earlier figures. Further n-fold filter-limiter combinations can be repeated at nodes 146-150 as previously described for the purposes of spectrum control.

Another embodiment of the present invention provides a mechanism for substantially removing all AM to PM conversion occurring within the processing path. Referring to FIG. 10, the input signal f_s is applied both to a first summation unit 270 and a down converter 272 which creates a second signal having reversed phase modulation with respect to the input but the same amplitude modulation. At the same time, the second signal is displaced from the idle frequency by a frequency of $-P$. Clearly, another mechanism could be used to generate the second signal, e.g. the two signals could be generated directly from a modulation source.

An idle frequency generator 274, arranged to produce frequency references at f_i and $2f_i$, applies the idle frequency f_i to the first summation unit 270 that is responsive to the input signal f_s. An output from the first summation unit 270 is applied to a first limiter 276, which first limiter produces some undesired AM to PM conversion, σ, in a constrained first zone signal emanating therefrom. To generate a first working frequency 281, an output signal 278 from the first limiter 276 is shifted upwards by an up converter 280 to a centre frequency of $2f_i$. The up converter 280 is therefore responsive to the idle frequency f_i output from idle frequency generator 274. The use of the up converter 280 merely corrects for an unavoidable frequency shift at other points in the processing path and ensures that the final wanted signal will be coherent with the input signal f_s. Signal 281 has a constant envelope and contains complex phase modulation α, as shown in FIG. 3(a) and as specified by eqn. 1. Signal 281 also has an undesired AM to PM term σ.

The down converter 272 receives, as a reference frequency from the idle frequency generator 274, twice the idle frequency, i.e., $2f_i$. The output from the

down converter 272 is such that signal 282 falls at a frequency $2f_i$-f_s (a difference of 2*P* from the true input signal). Any phase modulation has the opposite sign to that of the input signal f_s while the envelope is unaltered. Signal 282 is now summed with the idle frequency f_i in second summation unit 284 before application to a second limiter 286, identical in operation to the first limiter 276. A second output signal 288 (labelled in the diagram as "intermediate signal") emanating from the second limiter 286 has identical AM to PM conversion to that of the first limiter 276 although the phase modulation now has the opposite sign, i.e., $-\alpha$. Signal 288 is centred on f_i. Frequency converter 290 is arranged such that the difference frequency between signals 281 and 288 is taken. Moreover, the frequency converter is shown here for convenience as a parallel arrangement of an up-converter and a down converter, with the up-converter providing a sum signal and the down converter providing a difference (i.e. the neoteric) signal. As a result the neoteric signal centred at frequency f_i is formed, with phase modulation on signals 281 and 288 adding to give the desired phase modulation 2α while the spurious AM to PM conversion terms subtract to zero.

The basic architecture of FIG. 10, as previously described, can also function in an alternative way in which the up converter (in the frequency converter 290) takes the sum frequency of signals 281 and 288 to yield signal 296. Remembering that signals 281 and 288 have the same phase modulation α (but with an opposite sign,) and the same AM to PM conversion σ (but with a similar sign), the phase modulation will subtract to zero and the AM to PM conversion will add to 2σ. The result will be a third intermediate signal 296 centred on a frequency $3f_i$ containing the AM to PM conversion term alone.

Signal 288 passes to a frequency doubler 292, which generates the neoteric signal 293 containing phase modulation 2α at a centre frequency $2f_i$ as already discussed. However, in this case, the neoteric signal 293 also contains AM to PM conversion 2σ. Down converter 294 now takes the frequency difference

between signals 293 and 296. The result is the neoteric signal centred on a frequency f_i with the spurious AM to PM conversion 2σ subtracted out. It will be appreciated that for optimal operation of FIG. 10, the two paths should be substantially identical in their transfer responses (including absolute time delays, signal phases and filtering characteristics). In certain circumstances, e.g. where for some reason the input signal f_s has its phase modulation reversed, it may be advantageous to interchange the constrained first zone signal with the intermediate signal 288; the result then being the neoteric signal containing the input signal f_s with the correct phase. It will also be appreciated that the particular arrangement of the various frequency conversions is for the purpose of explanation only and that any arrangement whereby the AM to PM conversion term is subtracted from the neoteric signal may be equally used. Of course, n-fold filtering-limiting can also be applied, as previously described.

FIG. 10 also serves to demonstrate another mechanism whereby the neoteric signal can be generated, while it will be appreciated that the features of FIG. 8 (in relation to the injection of the signal and the idle frequency to correct for distortion to the wanted signal) can be imported into FIG. 10 on either one or both paths, as will be readily understood and notwithstanding that the configuration of FIG. 10 is substantially free of distortion. Indeed, the controlled adjustment of such injected signals can be used to counteract distortions occurring elsewhere in the neoteric signal path, e.g., those caused by band-limiting or unspecified effects within the amplifier to which the neoteric signal is applied.

It will be appreciated that the wanted signal is a replica of the input signal provided that its envelope "a" is always less than the amplitude of the idle frequency "b". If their ratio "r" exceeds unity then overload occurs and the result is, in general, undesirable. A choice of "b" much greater than the envelope will prevent such occurrence but then the neoteric signal would only contain a small wanted signal. Two methods to improve this situation suggest themselves,

namely: i) "b" can be varied at some slow rate to take account of changes in the envelope, e.g., it could follow syllabic variation in a speech signal; or ii) "b" can follow the envelope instantaneously, e.g. by assessing the envelope a soft limiting characteristic can be built into the present invention.

FIG. 11 illustrates another embodiment and is again based on FIGs. 5 and 6, although it is equally applicable, in principle, to FIGs. 7 to 10 and FIG. 12 to be described subsequently. Specifically, FIG. 11 contains a series combination of a summing unit 301, a limiter 304, a frequency doubler 306 and a frequency converter 308 for producing an neoteric signal 310 containing a wanted signal coherent with the input signal f_s. In addition, FIG. 11 contains an envelope detector 314 responsive to the input signal f_s, which envelope detector 314 operates to determine the instantaneous envelope of the incident signal, as will be readily understood. The output of the envelope detector 314 is quantitative and is processed according to a control algorithm 316 designed to achieve the objective of overload prevention. An algorithm output is used by level control unit 318 that accurately and instantaneously attenuates (modifies) the level "b" of the idle frequency. Since the neoteric signal is responsive only to the ratio a/b, the effect of the level control unit 318 is to change the absolute level of the wanted signal contained within the neoteric signal 310. Where such changes are made at the same rate as the changes in the envelope "a" of f_s, then (dependent upon the algorithm employed) absolute linearity will be lost but complementary benefits achieved in relation to some other associated parameter, e.g. an increase in the mean power of a variable envelope signal at the expense of an acceptable loss of linearity (sometimes called soft limiting).

Alternatively, an external level selector 320 may be used to define the level of the wanted signal to a precision governed only by the accuracy of the level control 318. In fact, an increase of *x* decibels in b will give a decrease of x decibels in the wanted signal. The level selector may, in fact, be switched into the level control unit 318 at the expense of the signal tap through the series

combination of the envelope detector and control algorithm.

An application to a more general type of signal path (containing both the internal path and the amplifier type of path) will now be described with particular reference to a receiver. It is common practice in state of the art receiver technology to delay any frequency selectivity until a late stage in the processing chain. Thus, much of the signal path in the receiver has a wide bandwidth in which many independent signals are present. The independent signals ostensibly combine to give a composite signal which has a much higher peak amplitude than any individual signal, while at the same time the composite signal has a large peak to mean power ratio (PMR). Any non-linear effects within the processing chain will generate a large number of intermodulation products, many of which fall in-band and can be harmful. Such effects occur in frequency converters and in wide- band amplifiers. In principle, the neoteric signal can improve reception; a) because, as already discussed, non-linearities up to the point where the neoteric signal has been formed are of no consequence (i.e., this is the internal path already discussed); and b) once the neoteric signal has been formed it can pass through an indefinite number of non-linear circuits and yet the wanted signal (a plurality of signals in this case) contained within it will remain undistorted.

An exemplary receiver/signal processing chain 300 that employs the underlying concepts of the present invention, namely the use of the neoteric signal to linearize a generalised path that includes a plurality of non-linear devices, is shown in FIG. 12. Incident signals 302 are received on an antenna 304. Typically, each of the many incident signals 302 will be modulated carriers and as such must be recovered, firstly by selecting them (filtering) and ultimately by demodulation. Incident signals 302 are applied to a summation unit 306 that also receives, as a second input, the idle frequency f_i (or a level modified version thereof) from idle frequency generator 308. A combined signal 310 is subsequently applied to a wide band low noise pre-amplifier 312. Output signals

can also be non-linear in respect to its transfer function. Reference frequency f_{ref} acts to tune the receiver 300 in a conventional manner by ensuring that the single desired signal for demodulation falls within the pass-band of narrowband amplifier 332. Signals output from frequency converter 330 are therefore effectively filtered by narrowband amplifier 332 to extract out a single desired signal 334 that will be present in the wanted signal (at the output of frequency converter 320) that is itself contained within the neoteric signal. The desired signal 334 can then be demodulated by conventional means in detector/demodulator 336.

As in the previous embodiments (such as FIGs. 5 and 6), the order of processing is unimportant, e.g., the hard limiter 324 and frequency doubler 326 could precede frequency converter 314, subject only to obtaining the correct neoteric signal.

Recovery of the wanted signal (generally) from the amplifier output or internal path will now be detailed. As has already been discussed, the neoteric signal provides a constant excitation for an amplifier (or for that matter, any signal path, even one with changes of frequency). The response contains the wanted signal that must be extracted and a set of unwanted components that must be rejected without harmful effects to the environment at large. Generally, at low powers, all that is required to recover the wanted signal is a bandpass filter, while at higher power levels it may be necessary for the various unwanted components to be absorbed in a load. Suitable architectures for recovery of the wanted signal from the incident neoteric signal are shown in FIGs. 13 and 14, which architectures are particularly applicable to relatively high power applications. In FIG. 13, the input signal f_s is signal processed (in block 350) to generate the neoteric signal 64 for application to an amplifier or signal path 354; this stage is representative of any one of FIGs. 5 to 10, for example. After communication of the neoteric signal through the amplifier, recovery of the wanted signal occurs in a bandpass filter 356 that is coupled to the signal path

354 via a circulator 358. Unwanted components that are reflected from the bandpass filter 356 are diverted to a load 360, such that power contained in the unwanted components is isolated and dissipated. For low to medium incident powers, the circulator 358 could be removed or replaced by an isolator and the load 360 could then be removed. It is also contemplated that for higher RF powers, for example, the energy reflected into the load 360 can be rectified and used to provide part of the power supply for the device in which the present invention is used. In this way, an amplifier, for example, may be partially driven by the transmitted envelope such that operation of a communication device becomes more energy efficient and use of the present invention more attractive.

In FIG. 14, two distinct but identical signal processors 350 are used with a shared idle frequency f_i, albeit that there is a phase shift of 180 degrees between the two. Consequently, neoteric signals 64 and 64^{bis} so formed will have the wanted signal contained within each neoteric signal in phase opposition. Neoteric signals 64 and 64^{bis} form the excitation for two substantially identical amplifiers 354. Respective amplifier outputs are then combined in hybrid 370 via port numbers two and four. An output from port number one of the hybrid 370 is a combined wanted signal 372 (obtained each amplifier) plus some residual parts of the unwanted component, namely those centred on $2f_i$-f_s, $4f_i$-$3f_s$ and so on. An output obtained from port number three 374 of the hybrid 370 contains the remainder of the unwanted components, namely those centred at f_i, $3f_i$-$2f_s$, $5f_i$-$4f_s$ and so on. The division of the unwanted components between the two ports occurs as a result of their respective signs, as determined in equation (2) above. The output from port number three is terminated in load 360. The combined wanted signal 372 (output from port number one) passes directly to band-pass filter 356 to select the wanted signal as before. Any residual unwanted components are reflected and finally appear at the output of amplifiers 354, where, however, their effects are assumed non-harmful. As already explained, the largest unwanted component falls at f_i and this is absorbed by the load. In this embodiment, the spectral components centred on f_i

are cancelled and can, in principle, fall within the signal band centred on f_i. Thus the spacing *P* between f_i and f_s can be reduced such that the spectrum of the unwanted component centred on $2f_i$-f_s does not overlap the wanted signal spectrum. As discussed in relation to FIG. 4b, this occurs for *P*>B.

The embodiment of FIG. 14 leads to a more relaxed filter requirement since the unwanted component that is closest in frequency to the wanted signal is absorbed in the load rather than appearing at the input of the bandpass filter 356. The neoteric signal bandwidth may also be reduced by suitable choice of *P*, as already discussed.

In summary, the neoteric signal is based upon a phase angle α between an input signal and a reference idle frequency with a frequency difference *P* as illustrated in FIG. 3. The neoteric signal is essentially a constant amplitude carrier carrying a phase modulation of substantially 2α. The neoteric signal contains two easily separable components, a wanted signal which is a replica of the input signal, and an unwanted component (which is itself the sum of other components in the spectrum domain). The choice of *P* determines the ultimate bandwidth of the neoteric signal and the ease of separating the wanted signal by filters or by subtraction.

The method of obtaining substantially double the phase angle is unimportant, although various methods employing, for example, direct generation of a second zone signal with frequency changes as necessary or purely digital processing have been illustrated.

The generation of the neoteric signal by means of practical devices is not perfect and residual distortion to the wanted signal will occur both within the processing to form the neoteric signal and also in the passage of this signal through an amplifier. Therefore the invention allows for small corrections (i.e., beneficial distortion) to be added to the neoteric signal. Generation of the

neoteric signal can be also nearly perfect by a method whereby undesired distortion caused by AM to PM conversion can be substantially cancelled. A purely digital generation can also yield a near perfect version of the neoteric signal.

The neoteric signal has two particularly desirable uses, namely it can drive a non-linear amplifier (conveniently expressed here as an external path) or a non-linear path can be made internal to the processing path through its use. The frequency band of the wanted or, output, signal need not be in the same frequency band as the input signal. The internal path has the merit that its bandwidth can be much smaller than the external path that uses the neoteric signal. In either case, the preferred embodiment of the present invention provides an amplification scheme that has an attainable peak power output substantially equal to the saturated output power. Beneficially, the preferred embodiment of the present invention does not require any adjustments to conventional amplifier configurations, but provides increased linearity to the saturation point as a straight line function while also avoiding complex processing of the signal envelope. The preferred embodiment of the present invention has application across the frequency spectrum. Advantageously, the level of the wanted signal contained within the neoteric signal generated by the preferred embodiment can be controlled, at a near instantaneous rate, to accrue instantaneous benefit (so called "soft overload") while at the same time minimizing harmful effects.

The principle of the present invention is generally applicable to any radio or cable system, but is particularly relevant to radio frequency communication systems and television (especially in relation to multi-carrier systems, such as coded orthogonal frequency division multiplex (COFDM)). Indeed, use of a constant amplitude by the present invention supports direct transmission methods through passive but non-linear transmission mediums. Such a medium might consist of an ionized gas or high-level ultra-sound.

Amplifiers utilizing the linearization technique of the present invention have application in linear repeaters that function to relay telecommunication signals between transmitting and receiving devices, such as a base station and a mobile cellular subscriber terminal. Furthermore, the present invention can be applied to cable distribution or optical networks either by making the cable distribution path or an optical modulator part of the internal path or by using the neoteric signal directly. Additionally, the present invention can be used within optical amplifiers to the extent that coherent optical sources can be manipulated like RF carriers since the present invention has the desirable quality of being able to linearize any signal path. In relation to their application in the field of optics, the preferred embodiments of the present invention can be practiced in optical amplifiers, for example. Specifically, the skilled addressee readily understands that light sources can be manipulated and combined to provide frequency doubling, while the functions of limiting and filtering are commonly practiced optical techniques. Therefore, provided that both the photon rate is much greater than the bandwidth of the neoteric signal and the light sources are reasonably coherent then the present invention may be used to linearise laser amplifiers or a non-linear optical path in the manner described above.

As will now be appreciated, the present invention is easily realized, namely through signal processing in a suitably configured microprocessor or digital signal processor (DSP) or even via analog phase modulation methods. Filters subsequently used to improve the clarity of the neoteric signal spectrum may be realized as either a discrete device or within the digital domain. A non-critical discrete filter is then able to recover the original excitation signal. The present invention has application in intermediately located amplifiers to support amplified outputs ranging from microwatts (mW) to megawatts (MW), although the present invention can be applied to any signal path existing between the processor and the recovering filter. Furthermore, the linearization process is not frequency dependent and so an output frequency need not be the same as the input

frequency. For example, the processor of the preferred embodiment of the present invention could be employed in a front end of a radio transceiver where up-conversion and down-conversion are required, respectively, for attaining a transmission frequency or an intermediate frequency for subsequent decoding.

It will, of course, be understood that the above description has been given by way of example only and that modifications in detail may be made within the scope of the present invention. It is emphasized that while use of the second zone of a hard limited signal represents a relatively simple way of generating a signal with the required modulation, it is envisaged that the underlying inventive concept can utilize varying angles having an angular displacement of something other than 2α relative to the idle tone f_i. In the particular analog embodiments chosen to illustrate the principles of the present invention (e.g. frequency multiplication), it is not considered particularly practicable to generate any arbitrary multiple of α by frequency multiplication, only integers. However, in a digital realization (or other analog realization in which the phase modulation is controllable), any multiple of α is possible and a small angular deviation δ from 2α still results in a meaningful linearity associated with the signal, although small non-linearities have now been introduced. Provided that δ is small the effect of using $2\alpha\pm\delta$ as an approximation merely adds distortion to the wanted signal, which distortion could nonetheless be used to correct for small distortions elsewhere in the path. Otherwise, there is no other advantage with using anything other than 2α. In other words, the value of the multiplier of α used to obtain the neoteric signal of the preferred embodiment is nominally numerically equal to two. Indeed, a digitally generated neoteric signal is preferable since imperfections (such as AM to PM conversion) are avoided leaving only residual imperfections, e.g. those created by narrowband filtering, to be removed by the control of δ.

While the present invention is principally described in relation to an information-bearing signal, it is emphasized that the underlying concept of generating the

neoteric (i.e. the double phase angle) signal is equally applicable to the precision control of any RF or optical signal and that, as such, the term "information-bearing signal" should be construed with the necessary latitude.

Realisation of the circuitry of the present invention is not limited to any particular technology, as will be readily appreciated, although the various functional elements may be suitably implemented within a semiconductor device, for example.

As indicated, it is considered necessary to elaborate and define the nomenclature that has been adopted above in relation to the technical description of the invention and to this effect the term:

- "amplifier" relates to any path to which a processed signal of the present invention (i.e. the "neoteric signal") is applied. As such, the path need not contain an amplification function, nor need its output be at the same frequency as its input. Indeed the path can be entirely passive, although the path usually corrupts a conventional signal in some way;
- "carrier" relates to a signal that supports the carriage of information;
- "central frequency" refers to the frequency of the information-bearing signal or carrier which acts as a reference from which any phase modulation is specified. It is also the frequency in the absence of such modulation;
- "distortion" applies to any undesirable imperfections caused to the "wanted signal";
- "envelope" is an imaginary line that defines a limit in the variation in signal level excursions between successive peaks of individual sine waves;
- "excitation" refers to the input signal to the principal amplifier;
- "frequency (up/down) converter" is a device for changing frequency and which uses a reference signal f_1 (usually emanating from a local oscillator) to convert a signal at frequency f_2 to a new frequency $f_1 \pm f_2$. An Up-converter (by convention) therefore adds the reference and signal frequencies, while a down converter takes the difference;

- "intermodulation" relates to spurious frequencies generated when several sine wave carriers of frequency f_n pass through a non-linear transfer function, specifically being of the form $lf_1+mf_1+nf_2+\ldots$ where l, m, n,... are positive or negative integers;
- "internal path" is a path within the processing function employed to form the combined signal of the preferred embodiment of the present invention. Specifically, when the term is used, it is taken to mean the second method by which the neoteric signal can render a path linear;
- "input signal" (sometimes interchangeably referred to as the "information-bearing signal") is the source signal that is to undergo processing to become the neoteric signal. The same signal also forms a part of the neoteric signal. The term is quite general and can be taken to include a plurality of separate uncorrelated signals. The signal or signals always meet the definition of narrowband;
- "limiter" refers to a commonly known device typically used in frequency modulation (FM) receivers. The limiter takes a narrowband signal with a variable envelope and constrains this signal to have a constant envelope. In principle the signal phase is unaltered. A logic comparator is illustrative of an analog limiter since it has a hard limited +V output for all positive inputs and -V for all negative inputs. When such a comparator is followed by a bandpass filter to extract its first zone the resulting signal will have a constant level. In relation to a non-linear amplifier, a limiter may, in practice, introduce undesired phase changes in the signal envelope;
- "narrowband" is used in relation to a signal that occupies a frequency band from f_{min} to f_{max} that can be completely described by the real part of its frequency spectrum. In a narrowband signal, there is no DC component and there is, in general, no restriction on the lower frequency limit f_{min}. Such narrowband signals can be accurately represented by a sine wave at a convenient centre frequency, which sine wave will have a modulated amplitude and argument (i.e., a modulated phase). In general, the steady or dc phase term is unimportant. Also, for the purpose of this application, it will